

图隐私项目周报

进度总结

1. 设计了一套基于K匿名的图数据隐私保护流程

因为种种限制（主要集中在算法层面），我们花费了较多的时间和精力不断迭代更新设计流程上，目前确定的方案为：

1) 使用基于k匿名的解匿名算法进行隐私风险的检测

目前设计的检测方法包括有基于度数、中心指纹、子图结构的解匿名法；此处用户可以设定多个隐私风险条件，这些条件之间可以存在一定的优先级，即设定哪些条件需要重点解决，哪些条件是次要条件。这样设定条件的好处我们可以先解决重点条件，在解决次要条件，降低运算压力的同时，关注更多的问题。

2) 可视化图中隐私问题

可视化图数据；使用bubbleSet可视化图隐私暴露风险节点；

3) 用户设定各种Utility属性

用户自由的添加系统内置的Utility属性(Degree, Joint Degree, Path length等)；也可以自定义添加Utility属性*；除了选择不同的Utility外，用户还可以设定不同Utility的比重，最终得到一个全局的Utility；

4) 对于1) 中检测出来为满足k匿名的局部问题提供k匿名解决方案推荐

这一步是关键步骤，为了保证系统的可用性，我们要尽可能的保证用户每一次选择的方案都能够消除数据中的部分隐私问题而不产生新的隐私问题，使得图的隐私风险逐步降低。经过我们对目前已有的匿名算法的调研发现，目前的匿名算法一般都是针对某种条件（例如，要保证Degree变化最小），对全局进行匿名化，即输入原始图，得到匿名化的结果图。这样一来，每种方案都有自身的侧重点，以致于多种方案之间会产生较多的冲突。如果系统选择这些全局方案，那么每一步采用的方案虽然解决了一定的隐私问题，但是很有可能在其他方面带来更多的隐私问题，不能保证隐私风险逐步降低的要求。因此我们希望，那些处于隐私暴露风险边界的节点在下一步中不会被处理，我正试图从目前已有的算法中，提取只针对局部问题，且能够考虑处于隐私暴露风险边界的节点的可用部分，手工实现特殊的匿名算法，保证流程的顺利进行。

5) 用户交互式地选择多种推荐的解决方案，比较方案之间Utility属性变化

针对用户关注的局部条件，我们给出一些推荐方案，每个局部问题的推荐方案之间都是针对当前图数据独立完成的，因此可能会存在冲突关系，我们将会将方案的推荐程度和方案之间的冲突关系使用可视化的形式表现出来；用户此处可以按照方案的推荐程度，选择互不冲突的方案加入到比较列表，进行Utility的比较。

6) 确定解决方案，可视化处理结果

经过比较，用户可以确定当前步骤需要执行的推荐方案，我们将通过可视化视图可视化当前结果图和原始图的差异（主要是边的增减）。

7) 循环1) ~6)

循环以上过程，直到用户能够得到满意的结果。

2. 算法收集

已拿到主流的7种Anonymization算法，9种De-Anonymization算法和22种Utility算法的java实现代码；但是由于上文中提到，这些匿名算法都是针对某种条件，对全局进行匿名化，这些全局算法都有一定的侧重点，单纯使用的话，会导致和其他方案之间的冲突，无法保证隐私风险的逐步下降，例如k-Degree方案保护了度数，但是Random Walk又会破坏k-Degree的结果，此外，运用多个全局算法将使得Utility大幅下降。目前，我们正尝试从这些算法中，抽取可用的部分，同时关注敏感节点（处于隐私风险边缘的节点），能够逐步消除隐私风险的同时，使得Utility不会下降太多。

3. 数据收集

目前拿到了网易游戏的玩家互动数据；

进度安排

• ~7.12

确定系统流程、De-anonymization条件和算法、Anonymization算法，可视化设计等，力图从整体和细节上跑通整个流程；

• 7.12 ~ 8.1

算法实现&界面设计；

• 8.1 ~

搭建Web系统；

下周工作

上周的任务安排为：

• 理解算法

叙萌、会华、如晟、致昊每人负责一篇匿名算法，进行理解，并思考：

1. 该方法是否为全局方法？若是，如果拆分成局部方法？(例如，对于k-degree方法，是否可以只针对degree==5的节点进行处理？)
2. 该算法是否为一个优化算法？是否可以通过该算法，推荐出若干解决方案？

• 试跑目前已有的算法代码

文龙负责这块的任务，并解决一下网易游戏互动数据中的隐私问题；

我们下周将针对以上两点工作进行讨论，期望结果：能够完成De-Anonymization和Anonymization的算法设计。

